Report

Current Biology

Deciphering Primordial Cyanobacterial Genome Functions from Protein Network Analysis

Highlights

- Network analysis elucidates ancient connections between different redox lifestyles
- Cyanobacteria and obligate anaerobes share the most core gene families
- Ancestral cyanobacterial functions include photosynthesis, hydrogenase, and defense
- Network analysis allows exploration of RC origin and evolution in bacterial lineages

Authors

Arye Harel, Slim Karkar, ...,
Paul G. Falkowski,
Debashish Bhattacharya

Correspondence

harelarik@gmail.com (A.H.),
debash.bhattacharya@gmail.com (D.B.)

In Brief

The origin of oxygenic cyanobacteria, over 2.5 billion years ago, is poorly understood. Harel et al. use network analysis to elucidate the position of cyanobacteria as an evolutionary bridge between anaerobic and obligate aerobic prokaryotes. They identify ancient cyanobacterial genome traits and explore the history of photosynthesis in microbes.

Harel et al., 2015, Current Biology *25*, 628–634
March 2, 2015 ©2015 Elsevier Ltd All rights reserved
http://dx.doi.org/10.1016/j.cub.2014.12.061

CellPress

Current Biology 25, 628–634, March 2, 2015 ©2015 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.12.061

Report

Deciphering Primordial Cyanobacterial Genome Functions from Protein Network Analysis

Arye Harel,[1,3,4,*] Slim Karkar,[3,4] Shu Cheng,[3] Paul G. Falkowski,[1,2] and Debashish Bhattacharya[3,*]
[1]Environmental Biophysics and Molecular Ecology Program, Department of Marine and Coastal Sciences, Rutgers, The State University of New Jersey, 71 Dudley Road, New Brunswick, NJ, 08901, USA
[2]Department of Earth and Planetary Sciences, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[3]Department of Ecology, Evolution, and Natural Resources, Rutgers, The State University of New Jersey, 59 Dudley Road, New Brunswick, NJ 08901, USA

Summary

The Great Oxidation Event (GOE) ~2.4 billion years ago resulted from the accumulation of oxygen by the ancestors of cyanobacteria [1–3]. Cyanobacteria continue to play a significant role in primary production [4] and in regulating the global marine and limnic nitrogen cycles [5, 6]. Relatively little is known, however, about the evolutionary history and gene content of primordial cyanobacteria [7, 8]. To address these issues, we used protein similarity networks [9], containing proteomes from 48 cyanobacteria as the test group, and reference proteomes from 84 microbes representing four distinct metabolic groups from most reducing to most oxidizing: methanogens, obligate anaerobes (nonmethanogenic), facultative aerobes, and obligate aerobes. These four metabolic groups represent extant bioinformatic proxies for ancient redox chemistries, extending from an anoxic origin through the GOE and ultimately to obligate aerobes [10–13]. Analysis of the network metric degree showed a strong relationship between cyanobacteria and obligate anaerobes, from which cyanobacteria presumably arose, for core functions that include translation, photosynthesis, energy conservation, and environmental interactions. These data were used to reconstruct primordial functions in cyanobacteria that included nine gene families involved in photosynthesis, hydrogenases, and proteins involved in defense from environmental stress. The presence of 60% of these genes in both reaction center I (RC-I) and RC-II-type bacteria may be explained by selective loss of either RC in the evolutionary history of some photosynthetic lineages. Finally, the network reveals that cyanobacteria occupy a unique position among prokaryotes as a hub between anaerobes and obligate aerobes.

Results and Discussion

Network Analysis Uncovers Evolutionary Connections between Cyanobacteria and Obligatory Anaerobes

We classified the genomes of 132 prokaryotes into five groups (cyanobacteria, C; archaeal methanogens, M; nonmethanogenic obligate anaerobes, OAN; facultative aerobes, FA; and obligate aerobes, OA; see Supplemental Experimental Procedures, Figure S1, and Table S1). Apart from cyanobacteria, the chosen lineages exhibit a wide range of redox capacities, with terminal electron acceptors that range from CO_2 to O_2, and provide proxies for widely different redox chemistries that were exploited by microbes during Earth's history and span the Great Oxidation Event (GOE) (e.g., methanogenesis to aerobiosis [10–12]). Using this metabolic classification system, we analyzed protein similarity networks [9, 14, 15] of 48 cyanobacteria with genomes from methanogens, obligate anaerobes, facultative aerobes, and obligate aerobe redox groups (see Supplemental Experimental Procedures). Analysis of the 132 proteomes in the network revealed conserved cyanobacterial core gene families (see Supplemental Experimental Procedures [10]) that are shared by at least 80% of the studied cyanobacteria. Usage of the four redox groups to analyze this core set shows that cyanobacteria share the largest number of unique core gene families with the obligate anaerobes (orange nodes in Figure 1; degree = 1, 83 core gene families, $p < 0.01$) followed by obligate aerobes (12 gene families, $p < 0.01$). The facultative aerobes and methanogens uniquely shared only 10 and 3 gene families, respectively, with cyanobacteria (Figure 1 and Table S4A). These results suggest either the retention of a significant ancient signal of an anaerobic ancestry in cyanobacteria or, alternatively, significantly higher horizontal gene transfer (HGT) between cyanobacteria and obligate anaerobes relative to the other redox groups (for details, see Table S4A).

Genome-wide Networks Recapitulate the Evolutionary Trajectory of Metabolic Groups

Determination of the number of protein families with degree = 1 when focusing on each group (cyanobacteria in Figure 2B, and the four redox groups in Figure 2C and Table S4) allowed us to generate a graph that represents overall network connectivity across the five groups. This graph (Figure 2A) shows the extent and putative direction of genetic information transfer between genomes of the four redox groups and cyanobacteria. The methanogens are clearly the most isolated (least connected) but share the most connections (thick red line) with obligate anaerobes. The next strongest connection is from obligate anaerobes to facultative aerobes and then from facultative aerobes to obligate aerobes. This pattern recapitulates the hypothesized direction of evolution of prokaryotes from methanogens and sulfate reducers that are deeply rooted in the tree of life [16–18] through iron reducers prevailing in the Archean eon [19, 20]. This was followed by utilization of novel electron acceptors that were formed as the Earth became increasingly oxidized (i.e., the formation of nitrate [21]), which ultimately led to the exploitation of the high oxidation potential of oxygen, and to the evolution of obligate aerobes. The position of cyanobacteria as being highly connected to two groups spanning the GOE (i.e., core gene families from obligate anaerobes, Figure 1, and the total set of gene families from obligate aerobes, Figure 2B and [10]) highlights the metabolic versatility [22] of cyanobacteria and the composite nature of their genomes that harbor core genes from anaerobes as well as many gene families from aerobes.

[4]Co-first author
*Correspondence: harelarik@gmail.com (A.H.), debash.bhattacharya@gmail.com (D.B.)

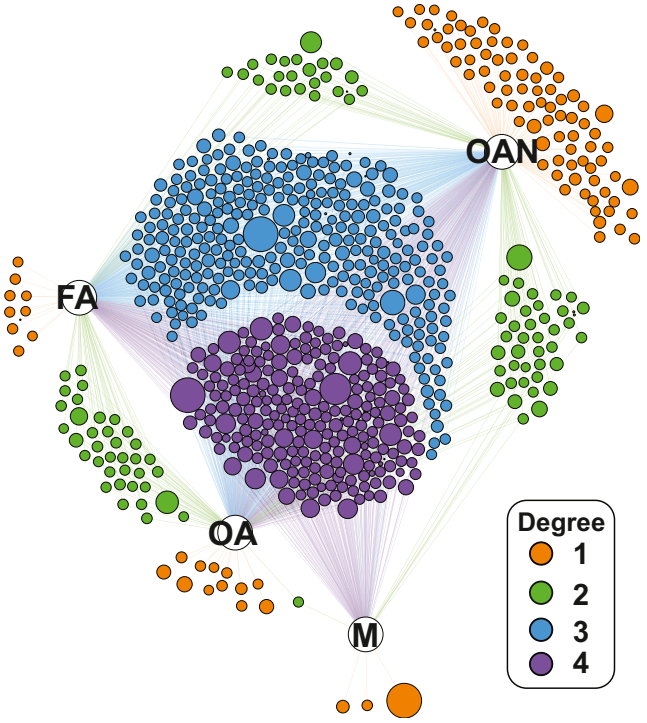

Figure 1. "Projected Network" of Cyanobacterial Gene Families and Their Relationships to Members of the Four Redox Groups

In this representation (see Supplemental Experimental Procedures), only core gene families from cyanobacteria (i.e., shared by 80% of these taxa) that have homologs in any of the four redox groups are shown. Degree = 1, shown with orange circles, denotes protein families that are shared uniquely by cyanobacteria (C) and one of archaeal methanogens (M), facultative aerobes (FA), obligate aerobes (OA), and nonmethanogenic obligate anaerobes (OAN), whereas degree = 4 indicates that C shares these gene families with all four redox groups (shown with purple circles; Table S1). Other degrees are interpreted in the same way. Note that C shares the most core gene families with the OAN group followed by the OA group. The location of the nodes reflects the qualitative composition of the gene families, in terms of homologs from noncyanobacterial groups (i.e., M, FA, OA, and OAN, each represented by an open, labeled node; see Supplemental Experimental Procedures and Table S1).

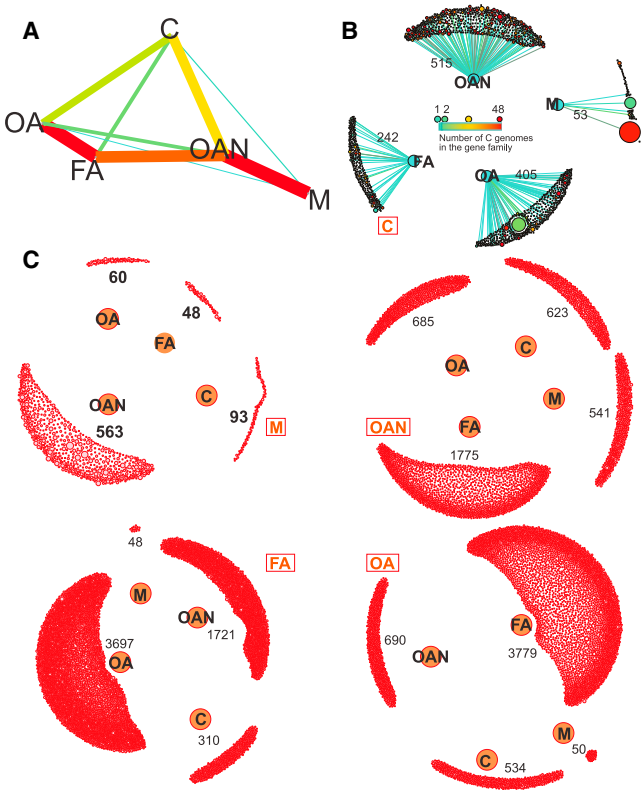

Figure 2. Summary of Gene Family Sharing between Cyanobacteria and the Four Redox Groups

(A) This network is obtained from the matrix of all (not just core) exclusively shared gene families (degree = 1; see Table S4B), with branch thicknesses indicating the extent of gene sharing and color denoting significance (decreasing significance from red to yellow for enrichment and light blue to green for depletion; see Table S4B). For details of the analysis used to generate (A), see Supplemental Experimental Procedures.

(B) This projected network shows C gene families that are shared exclusively with other groups (nodes of degree = 1; see Supplemental Experimental Procedures). Colors (green to yellow to red) indicate the number of cyanobacterial genomes represented. The C core gene families (present in >80% of these genomes, i.e., >39 genomes) are indicated in red; node sizes reflect the number of proteins that comprise the gene families. The numbers of gene families are reported in the C column of the matrix in Table S4B.

(C) These projected networks show M, OAN, FA, and OA gene families that are shared exclusively with other groups (nodes of degree = 1), also reported in Table S4B.

Analysis of Cyanobacterial Core Genes Shared with Obligate Anaerobes

Network analysis revealed gene families that link core cyanobacterial genes and obligatory anaerobes, with the exclusion of any other group (exclusively shared gene families of degree = 1 in Figure 1). Those genes encode functions that may have been essential for the common ancestor of cyanobacteria (hereafter denoted "pre-C"; see numerous core genes in Figure 2B, Table 1, and Table S2). It is generally challenging to obtain compelling evidence for such ancient signatures of shared history, in particular when considering the metabolic versatility of cyanobacteria [22]. Nonetheless, our results are statistically significant (Table S4 and Supplemental Experimental Procedures section "Statistical Significance of Exclusive Gene Families"; see also [10]) and are in line with the assumption that the pre-C lived in an anaerobic environment [1–3]. With roles in the regulation of protein synthesis, potentially correlated with environmental cues (see below), photosynthesis, energy conservation, and transport, these functions would presumably have been essential for survival of the pre-C.

Existing Scenarios for the Origin of Cyanobacteria and Oxygenic Photosynthesis

The origin and evolution of cyanobacteria is a subject of ongoing debate, with two competing hypotheses under consideration [23, 24]. The first scenario posits the existence of a pre-C that contained both type I and II reaction centers (RCs) that share a common ancestry through ancient gene duplications. These RCs were vertically inherited in this lineage and transmitted to anoxygenic photosynthetic bacteria that later underwent selective loss of either RC-I or RC-II [7, 25–28]. In the second scenario, cyanobacteria evolved from the fusion of two existing bacterial lineages that contained either RC-I or RC-II, or through HGT of one of the RCs to a lineage that already contained the other reaction center [2, 29–31]. The observation that the cyanobacterium *Oscillatoria limnetica* utilizes sulfide as a sole electron donor in photosystem I (PSI)-mediated

Table 1. Core Cyanobacterial Functions Associated with Obligate Anaerobes

KEGG Orthologs			RC-I			RC-II				
ID	Category	Short Description	Chl	Hel	GSB	GNSB	PSB	PNSB	Mel	Groups
Functions Found in Both RC-I and RC-II-Containing Bacteria										
13292	*environ. int.-membrane*	prolipoprotein diacylglycerol transferase	1	1	3	4	6	5	3	6
3629	*environ. int.-stress*	DNA replication and repair protein	1	1	4	4	2	3	3	6
3404	PHS	magnesium chelatase subunit D	1	1	4	4	5	4	0	6
4038	PHS	light-independent protochlorophyllide reductase subunit N	1	1	5	4	6	5	0	6
4039	PHS	light-independent protochlorophyllide reductase subunit B	1	1	5	4	6	5	0	6
3437	translation regulation	tRNA methytrans.	1	1	4	4	5	2	2	6
10960	PHS	geranylgeranyl reductase	1	0	5	4	6	5	0	5
441	PHS+resp.	*F420 hyd.* subunit B	1	0	3	3	6	4	0	5
2113	PHS+resp.	ATPase subunit delta	1	1	4	0	6	5	2	5
3215	translation regulation	23S rRNA u-methytrans.	1	0	4	4	6	4	2	5
1810	carbon metabolism	glucose-6-phosphate isomerase	0	0	1	2	4	3	0	4
8296	*environ. int.-sensing*	phosphohistidine phosphatase	1	0	0	1	3	1	1	4
4079	*environ. int.-stress*	HSP90A	0	0	2	4	6	3	0	4
9818	*environ. int.-transport*	Mg/Fe transport	0	0	1	4	2	4	0	4
9835	*PHS and environ int.-stress*	prolycopene isomerase	1	0	5	1	1	0	0	4
3500	translation regulation	16S rRNA c-methytrans.	1	0	4	0	6	5	1	4
966	carbon metabolism	mannose-1-phosphate guanylyltransferase	1	0	4	4	0	0	2	3
3568	*environ. int.-biotic*	TldD protein (microcin processing)	1	0	0	0	3	4	1	3
8481[a]	*environ. int.-stress*	circadian clock protein	0	0	0	4	4	2	0	3
3116	*environ. int.-transport*	translocase TatA	0	1	1	0	2	0	1	3
2293	*PHS and environ int.-stress*	15-cis-phytoene desaturase	1	0	5	2	0	0	0	3
3536	translation regulation	ribonuclease P protein	1	1	0	1	0	0	1	3
3218	translation regulation	23S rRNA-guanosine methytrans.	0	0	1	0	0	2	0	2
Functions Found Only in RC-I or RC-II-Containing Bacteria										
13821	amino acid metabolism	proline dehydrogenase/pyrroline-5-carboxylate dehydrogenase	0	0	0	0	1	2	0	2
1423	amino acid metabolism	E3.4.-.-	0	0	0	1	0	0	0	1
1583	amino acid metabolism	arginine decarboxylase	0	1	0	0	0	0	2	1
1758	amino acid metabolism	cystathionine gamma-lyase	0	1	0	0	0	0	2	1
07567[a]	amino acid metabolism	TdcF protein	0	0	0	0	1	0	0	1
13282	C and N metabolism	cyanophycinase	0	0	0	1	0	0	0	1
507	carbon metabolism	stearoyl-CoA desaturase	0	0	0	2	0	0	1	1
2077	environ. int.-transport	Zn/Mn transport	0	0	0	2	0	0	0	1
6147	environ. int.-transport	ATP-binding cassette	0	0	2	0	0	0	0	1
2635	PHS and resp. in cyanobacteria	cytochrome b6	0	1	0	0	0	0	0	1
7263	protein metabolism	Zn protease	0	0	0	0	1	0	0	1
1582	amino acid metabolism	lysine degradation	0	0	0	0	0	0	2	0
788	carbon metabolism	thiamine-P pyrophosphorylase	0	0	0	0	0	0	0	0
8884[a]	PTM	–	0	0	0	0	0	0	0	0

Functions encoded by connected components comprising solely proteins from cyanobacteria and obligate anaerobes (primarily OAN and some M, as indicated; Tables S2 and S3), and their distribution among genomes of six groups of photosynthetic prokaryotes (Chl, Chloracidobacterium; Hel, heliobacteria; GSB, green sulfur bacteria; GNSB, green nonsulfur bacteria; PSB, purple sulfur bacteria; PNSB, purple nonsulfur bacteria) and melainabacteria (Mel; see Supplemental Experimental Procedures). Italics indicate functions related to interactions with environmental cues. Data are sorted by number of photosynthetic groups (final column). For the KEGG ortholog IDs, the initial "k" was removed for simplicity. Categories are based on a literature review and KEGG annotations. Abbreviations: environ., environmental; hyd., hydrogenase; int., interactions; methytrans., methyltransferase; PHS, photosynthesis; PTM, posttranslation modification; resp., respiration.

[a]Protein family found in components containing only methanogens and cyanobacteria. All other entries indicate components that contain only OAN and cyanobacteria.

anoxygenic photosynthesis [32–35] putatively links photosynthetic anaerobes to primordial cyanobacteria. This electron donor was presumably abundant in primary producers from ancient anoxygenic environments [34], which further suggests that such cyanobacteria could serve as a missing link in the evolution of oxygenic photosynthesis. Comparative phylogenetic analysis of 13 cyanobacterial genomes revealed a core set of 323 genes (encoding components of the photosynthetic and ribosomal apparatus) with similar evolutionary histories [8]. In another study, a core set of 1,054 gene families was found by clustering orthologs in 15 cyanobacterial genomes [7]. Among these core genes, 936 families were present in other bacteria and were involved in DNA replication and repair, transcription, translation, key metabolic pathways, and energy metabolism. Only a handful of components of the core cyanobacterial genes that encode the photosynthetic apparatus (e.g., involved in bacteriochlorophyll synthesis) were found in anoxygenic photosynthetic bacteria [7].

Genetic Toolkit in the Pre-C

Our results identified nine protein families that participate in photosynthesis and are present in the postulated pre-C (Figure 3 and Table 1). Six of these components mediate the biosynthesis of photosynthetic pigments (chlorophylls and

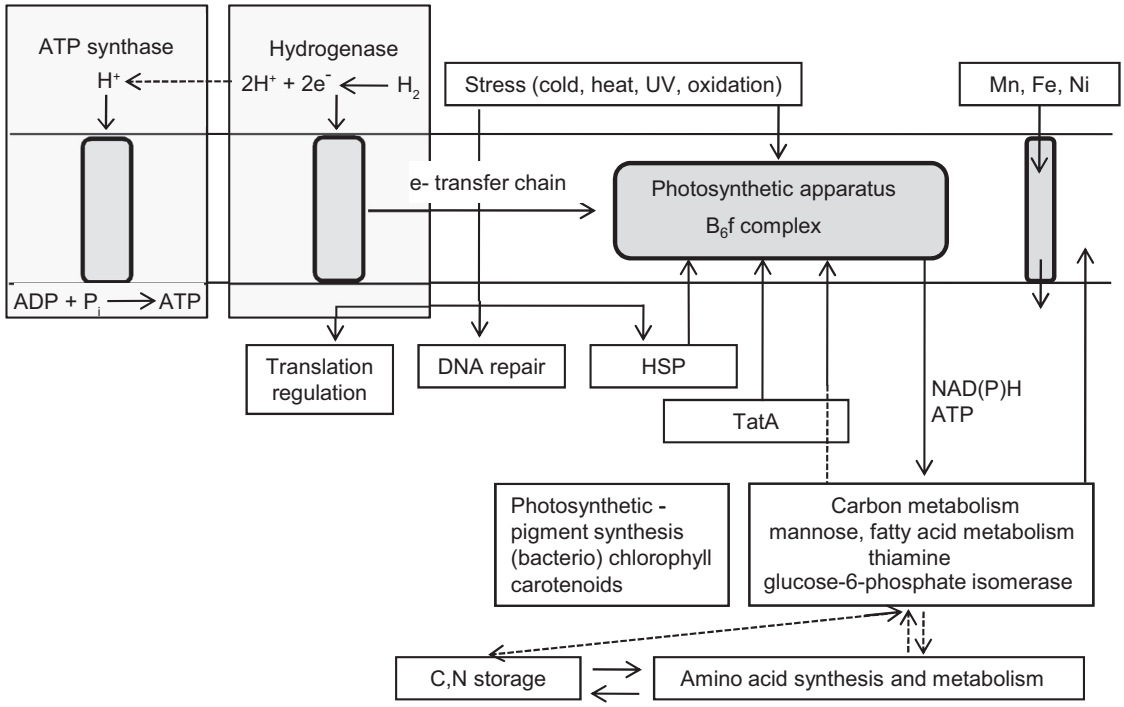

Figure 3. Reconstruction of Putative Metabolic Functions in the Pre-C

The components that comprise the core genome of cyanobacteria and obligatory anaerobes (bacteria or methanogens) include the following metabolic functions and therefore presumably represent ancestral functions: H_2 hydrogenase (KEGG ID 00441), transport of metal ions (KEGG IDs 02077, 09818, 06147), regulation of translation (KEGG IDs 03500, 03215, 03437, 03536, K03218), b_6f complex (KEGG ID 02635), ATP synthase (KEGG ID 02113), stress HSP90 (KEGG ID 004079), secretion TatA (KEGG ID 03116), DNA repair (KEGG ID 03629), sensing environmental cues (KEGG ID 08296), carotenoid biosynthesis (KEGG IDs 02293, 09835), (bacterio)chlorophyll synthesis (KEGG IDs 03404, 4038/9, 10960), amino acid synthesis and metabolism (KEGG IDs 01758, 01582, 01583, 01423, 13821, 13282, 13292, 03568, 09022), and carbon metabolism (KEGG IDs 00788, 01810, 00966, 00507). HSP stands for heat shock protein (see also Tables 1 and S2).

carotenoids). These include two subunits of the light-independent protochlorophyllide reductase, (B)ChlB and (B)ChlN. Absence of the third subunit, (B)ChlL, in the putative pre-C toolkit could be explained by higher sequence divergence of the (B)ChlL-based dimer when compared to the more structurally constrained (B)ChlB, (B)ChlN-based heterodimer [36]. Functions mediating the biosynthesis of photosynthetic pigments include representatives from *Heliobacterium modesticaldum* and the green sulfur bacterium (GSB) *Chlorobium limicola*. The common ancestry of these two groups (GSB and heliobacteria) and cyanobacteria is also supported by phylogenetic analyses of enzymes involved in (bacterio)chlorophyll biosynthesis [30] and Rieske/cytochrome *b* binding proteins [37, 38], which show them as sister clades or place GSB/heliobacteria as ancestors of cyanobacteria [30]. This potential evolutionary connection is further supported by the fact that these three groups (GSB, heliobacteria, and cyanobacteria) contain a RC-I [31]. We cannot rule out HGT from cyanobacteria to obligatory anaerobes to explain these results; however, our findings are consistent with analyses of components of the photosynthetic apparatus in core genomes of cyanobacteria [8] and orthologs of anoxygenic phototrophic bacteria [7]. The latter data led to the hypothesis that cyanobacteria (i.e., the pre-C) were derived from anoxygenic phototrophs harboring a RC-I-like reaction center (such as GSB and heliobacteria).

Another potential function that could have been of fundamental importance to the pre-C includes homologs of coenzyme F420 hydrogenase beta subunit (an F420 binding subunit, Figure 3, and KEGG ortholog K0441 in Tables 1 and S2). Phylogenetic analysis of its large and small subunits suggests that F420 hydrogenase (EC 1.12.99.1) is an early-diverging enzyme that used H_2 as the electron donor in energy-conserving electron transfer chains [39]. Although no functional form of this enzyme has yet been demonstrated in cyanobacteria (or bacteria), these taxa contain phylogenetically related oxygen-sensitive bidirectional NAD(P) hydrogenases capable of oxidizing H_2 [13, 39]. It was previously suggested, based on the abundance of minerals, that the primary electron donor of early photosynthetic autotrophic prokaryotes in microbial mats was probably H_2 [40, 41]. Hydrogenase-derived $H_2/2H^+$ reactions can supply sufficient reduction potential (E_m' −410 mV) to reduce intermediates like $NAD(P)^+$ (E_m' −320 mV), which are used for carbon fixation [41], making this reaction potentially valuable for the anoxygenic-photosynthetic pre-C. A handful of studies elucidated H_2-dependent anoxygenic photosynthesis in three different genera of cyanobacteria [42–44], including RC-I-mediated carbon photoassimilation [42], supporting the hypothesis that the pre-C harbored a homologous pathway. To the best of our knowledge, the hydrogenase genes mediating this process in cyanobacteria have not been identified; however, the finding that five of the six known bacterial photosynthetic groups (see below) harbor a related hydrogenase (Table 1) indicates that this enzyme has an ancient origin, potentially associated with photosynthesis. Our analysis reveals two additional key elements in the energy conservation pathway that are found in six anaerobic bacteria, including

heliobacteria: a F-ATPase H^+-transporting subunit delta of the conserved ATP synthase machinery, present in the last common ancestor of archaebacteria and eubacteria [24], and the b_6 protein of the cytochrome b_6f complex participating in the electron transfer chain of both photosynthesis and respiration in cyanobacteria [45].

Elements of the Ancient Genetic Toolkit of the Pre-C Present Both in RC-I- and RC-II-Harboring Photosynthetic Lineages

To gain insights into the evolution of cyanobacterial photosynthesis, we searched for orthologs that presumably existed in the pre-C. For this analysis, based on functions detected in the pre-C genetic toolkit, we included representatives from the six known groups of photosynthetic bacteria (see Supplemental Experimental Procedures and Tables 1, S2, and S3). The observation that most (90%) of the functions found in the pre-C are present in at least one of the photosynthetic groups is consistent with putative photosynthetic capability in the pre-C (Table 1). Nine of these functions are related to interactions with environmental cues (indicated in italics in Table 1), such as defense against abiotic oxidative stress (HSP90A [46]; carotenoids [11]) and UV stress (recF [47]), which would have been important under the intense UV radiation that presumably existed in the pre-GOE environment of the cyanobacterial ancestor [48]. These "peripheral" functions contrast with the fundamental metabolic machinery that must have originated prior to the split of the pre-C. Existing scenarios for the origin of the RC-I+RC-II-containing cyanobacteria are based on understanding the evolution of individual RC-I- or RC-II-harboring lineages [2, 7, 25, 27, 29, 31]. Our finding that 62% of core cyanobacterial functions are present in both RC-I- and RC-II-type lineages is consistent with the hypothesis that anoxygenic photosynthetic bacteria likely underwent selective loss of one of the two RCs [7, 25, 27]. Seven of these core functions are associated with photosynthesis, further supporting this hypothesis. However, other non-photosynthesis-related functions provide only indirect support for the selective loss of RC-related genes, because their presence in the pre-C core gene set and in both RC-I- and RC-II-containing lineages suggests that there was a common ancestor that contained both RCs [24]. One potential scenario is that an early pre-C utilized RC-I-based photosynthesis fueled by an alternative available electron donor, in the same manner that sulfide is utilized by *O. limnetica*. Considering their similarity in structure and function, it has been suggested that RC-II could have evolved later from duplication and divergence of RC-I [2, 7, 26, 49]. Finally, loss of either of the RCs from the presumed common pre-C ancestor and subsequent divergence would explain the distribution of RC in all photosynthetic lineages [24]. Such losses could potentially have been linked to extensive (multimillion-year) episodes of transient anoxic conditions interspersed with oxic conditions prior to the GOE [21, 50]. We cannot rule out that RC-I- and RC-II-harboring lineages may have acquired their reaction center via HGT of either one, from the pre-C over evolutionary time. Acquisition of RC-II before it was associated with the oxygen evolution complex may explain the absence of the latter in some of these lineages [7]. Nonetheless, the finding of selective loss of either RC is in line with a recent phylogenetic analysis of the core (bacterio) chlorophyll biosynthetic pathway that did not support the fusion model of RC origin because the division of chlorophyll biosynthesis genes into RC-I- and RC-II-type clades was not observed [28].

Finally, recent studies of environmental samples (i.e., human gut, groundwater, and wastewater) have revealed a deep-branching sister clade (i.e., melainabacteria, named after the Greek nymph of dark waters) of cyanobacteria that does not encode genes involved in photosynthesis [51, 52]. Based on genomic analysis (comparative and phylogenetic) and the absence of the photosynthetic machinery, it was hypothesized that melainabacteria and cyanobacteria form sister clades that shared a nonphotosynthetic ancestor [51, 52]. Our analysis of core cyanobacterial functions found in obligatory anaerobes reveals 15 functions present in three genomes from diverse clades of melainabacteria (see Supplemental Experimental Procedures and Table 1). Fourteen of these functions were detected in other photosynthetic groups, including five that participate in peripheral pathways. These results are more in line with the alternative hypothesis that melainabacteria and cyanobacteria likely shared a photosynthetic common ancestor. However, the existing "absence of evidence" data (i.e., components of the photosynthetic machinery are not present in available melainabacteria genomes) do not allow us to clearly distinguish between these two scenarios for pre-C evolution.

In summary, we recognize the great difficulties in reconstructing evolutionary relationships in deep time, and moreover in predicting the genomic capacity of ancient genomes. The utility of protein networks lies in their ability to uncover ancient signals in sequence data, providing a complementary approach to phylogenetics [10, 53]. Here we used network methods to gain insights into the rise of cyanobacteria in the context of the GOE. Our results identify a set of functions that appear to have been essential for a postulated anoxygenic photosynthetic ancestor of cyanobacteria (the pre-C) that used H_2 as an electron source and later "invented" the oxygen evolution complex to give rise to oxygenic photosynthesis [7, 8]. These innovations led to the rise of a lineage of oxygenic photosynthetic organisms that expanded into diverse niches in the biosphere and ultimately provided the foundation for the rise of algae and plants through plastid primary endosymbiosis [54].

Supplemental Information

Supplemental Information includes one figure, four tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.12.061.

Author Contributions

A.H. and D.B. conceived the study. A.H. and S.K. designed and conducted the bioinformatic analyses and wrote the initial manuscript draft. S.C. performed the preliminary bioinformatic analyses. P.G.F. and D.B. read and commented on the manuscript.

Acknowledgments

This research was funded by the Gordon and Betty Moore Foundation through grant GBMF2807 to P.G.F. We thank Huan Qiu and other D.B. lab members for helpful discussions.

Received: August 4, 2014
Revised: November 5, 2014
Accepted: December 29, 2014
Published: February 12, 2015

References

1. Bekker, A., Holland, H.D., Wang, P.L., Rumble, D., 3rd, Stein, H.J., Hannah, J.L., Coetzee, L.L., and Beukes, N.J. (2004). Dating the rise of atmospheric oxygen. Nature *427*, 117–120.

2. Blankenship, R.E. (1992). Origin and early evolution of photosynthesis. Photosynth. Res. *33*, 91–111.
3. Canfield, D.E. (2014). Oxygen: A Four Billion Year History (Princeton University Press).
4. Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. Nature *277*, 293–294.
5. Capone, D.G., Zehr, J.P., Paerl, H.W., Bergman, B., and Carpenter, E.J. (1997). *Trichodesmium*, a globally significant marine cyanobacterium. Science *276*, 1221–1229.
6. Zehr, J.P., Waterbury, J.B., Turner, P.J., Montoya, J.P., Omoregie, E., Steward, G.F., Hansen, A., and Karl, D.M. (2001). Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. Nature *412*, 635–638.
7. Mulkidjanian, A.Y., Koonin, E.V., Makarova, K.S., Mekhedov, S.L., Sorokin, A., Wolf, Y.I., Dufresne, A., Partensky, F., Burd, H., Kaznadzey, D., et al. (2006). The cyanobacterial genome core and the origin of photosynthesis. Proc. Natl. Acad. Sci. USA *103*, 13126–13131.
8. Shi, T., and Falkowski, P.G. (2008). Genome evolution in cyanobacteria: the stable core and the variable shell. Proc. Natl. Acad. Sci. USA *105*, 2510–2515.
9. Halary, S., Leigh, J.W., Cheaib, B., Lopez, P., and Bapteste, E. (2010). Network analyses structure genetic diversity in independent genetic worlds. Proc. Natl. Acad. Sci. USA *107*, 127–132.
10. Cheng, S., Karkar, S., Bapteste, E., Yee, N., Falkowski, P., and Bhattacharya, D. (2014). Sequence similarity network reveals the imprints of major diversification events in the evolution of microbial life. Front. Ecol. Evol. *2*, 72.
11. Madigan, M.T. (2012). Brock Biology of Microorganisms, Thirteenth Edition (Benjamin Cummings).
12. Martin, W.F., Sousa, F.L., and Lane, N. (2014). Evolution. Energy at life's origin. Science *344*, 1092–1093.
13. Vignais, P.M., and Billoud, B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. Chem. Rev. *107*, 4206–4272.
14. Bapteste, E., Lopez, P., Bouchard, F., Baquero, F., McInerney, J.O., and Burian, R.M. (2012). Evolutionary analyses of non-genealogical bonds produced by introgressive descent. Proc. Natl. Acad. Sci. USA *109*, 18266–18272.
15. Doolittle, W.F. (1999). Lateral genomics. Trends Cell Biol. *9*, M5–M8.
16. Shen, Y., Buick, R., and Canfield, D.E. (2001). Isotopic evidence for microbial sulphate reduction in the early Archaean era. Nature *410*, 77–81.
17. Slesarev, A.I., Mezhevaya, K.V., Makarova, K.S., Polushin, N.N., Shcherbinina, O.V., Shakhova, V.V., Belova, G.I., Aravind, L., Natale, D.A., Rogozin, I.B., et al. (2002). The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. Proc. Natl. Acad. Sci. USA *99*, 4644–4649.
18. Wagner, M., Roger, A.J., Flax, J.L., Brusseau, G.A., and Stahl, D.A. (1998). Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. J. Bacteriol. *180*, 2975–2982.
19. Heimann, A., Johnson, C.M., Beard, B.L., Valley, J.W., Roden, E.E., Spicuzza, M.J., and Beukes, N.J. (2010). Fe, C, and O isotope compositions of banded iron formation carbonates demonstrate a major role for dissimilatory iron reduction in ~2.5 Ga marine environments. Earth Planet. Sci. Lett. *294*, 8–18.
20. Johnson, C.M., Beard, B.L., Klein, C., Beukes, N.J., and Roden, E.E. (2008). Iron isotopes constrain biologic and abiologic processes in banded iron formation genesis. Geochim. Cosmochim. Acta *72*, 151–169.
21. Godfrey, L.V., and Falkowski, P.G. (2009). The cycling and redox state of nitrogen in the Archaean ocean. Nat. Geosci. *2*, 725–729.
22. Voorhies, A.A., Biddanda, B.A., Kendall, S.T., Jain, S., Marcus, D.N., Nold, S.C., Sheldon, N.D., and Dick, G.J. (2012). Cyanobacterial life at low O(2): community genomics and function reveal metabolic versatility and extremely low diversity in a Great Lakes sinkhole mat. Geobiology *10*, 250–267.
23. Allen, J.F., and Martin, W. (2007). Evolutionary biology: out of thin air. Nature *445*, 610–612.
24. Hohmann-Marriott, M.F., and Blankenship, R.E. (2011). Evolution of photosynthesis. Annu. Rev. Plant Biol. *62*, 515–548.
25. Allen, J.F. (2005). A redox switch hypothesis for the origin of two light reactions in photosynthesis. FEBS Lett. *579*, 963–968.
26. Baymann, F., Brugna, M., Mühlenhoff, U., and Nitschke, W. (2001). Daddy, where did (PS)I come from? Biochim. Biophys. Acta *1507*, 291–310.
27. Olson, J.M., and Pierson, B.K. (1987). Origin and evolution of photosynthetic reaction centers. Orig. Life Evol. Biosph. *17*, 419–430.
28. Sousa, F.L., Shavit-Grievink, L., Allen, J.F., and Martin, W.F. (2013). Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. Genome Biol. Evol. *5*, 200–216.
29. Nisbet, E.G., and Sleep, N.H. (2001). The habitat and nature of early life. Nature *409*, 1083–1091.
30. Xiong, J., Fischer, W.M., Inoue, K., Nakahara, M., and Bauer, C.E. (2000). Molecular evidence for the early evolution of photosynthesis. Science *289*, 1724–1730.
31. Xiong, J. (2006). Photosynthesis: what color was its origin? Genome Biol. *7*, 245.
32. Cohen, Y., Jørgensen, B.B., Padan, E., and Shilo, M. (1975). Sulphide-dependent anoxygenic photosynthesis in the cyanobacterium *Oscillatoria limnetica*. Nature *257*, 489–492.
33. Cohen, Y., Padan, E., and Shilo, M. (1975). Facultative anoxygenic photosynthesis in the cyanobacterium *Oscillatoria limnetica*. J. Bacteriol. *123*, 855–861.
34. Johnston, D.T., Wolfe-Simon, F., Pearson, A., and Knoll, A.H. (2009). Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. Proc. Natl. Acad. Sci. USA *106*, 16925–16929.
35. Oren, A., Padan, E., and Avron, M. (1977). Quantum yields for oxygenic and anoxygenic photosynthesis in the cyanobacterium *Oscillatoria limnetica*. Proc. Natl. Acad. Sci. USA *74*, 2152–2156.
36. Muraki, N., Nomata, J., Ebata, K., Mizoguchi, T., Shiba, T., Tamiaki, H., Kurisu, G., and Fujita, Y. (2010). X-ray crystal structure of the light-independent protochlorophyllide reductase. Nature *465*, 110–114.
37. Lebrun, E., Santini, J.M., Brugna, M., Ducluzeau, A.L., Ouchane, S., Schoepp-Cothenet, B., Baymann, F., and Nitschke, W. (2006). The Rieske protein: a case study on the pitfalls of multiple sequence alignments and phylogenetic reconstruction. Mol. Biol. Evol. *23*, 1180–1191.
38. Nitschke, W., van Lis, R., Schoepp-Cothenet, B., and Baymann, F. (2010). The "green" phylogenetic clade of Rieske/cytb complexes. Photosynth. Res. *104*, 347–355.
39. Boyd, E.S., Schut, G.J., Adams, M.W.W., and Peters, J.W. (2014). Hydrogen metabolism and the evolution of biological respiration. Microbiology *9*, 361–367.
40. Kim, J.D., Yee, N., Nanda, V., and Falkowski, P.G. (2013). Anoxic photochemical oxidation of siderite generates molecular hydrogen and iron oxides. Proc. Natl. Acad. Sci. USA *110*, 10073–10077.
41. Tice, M.M., and Lowe, D.R. (2006). Hydrogen-based carbon fixation in the earliest known photosynthetic organisms. Geology *34*, 37–40.
42. Belkin, S., and Padan, E. (1978). Hydrogen metabolism in the facultative anoxygenic cyanobacteria (blue-green algae) *Oscillatoria limnetica* and *Aphanothece halophytica*. Arch. Microbiol. *116*, 109–111.
43. Peschek, G.A. (1979). Anaerobic hydrogenase activity in *Anacystis nidulans*. H2-dependent photoreduction and related reactions. Biochim. Biophys. Acta *548*, 187–202.
44. Serebryakova, L., Novichkova, N., and Gogotov, I. (2002). Facultative H-2-dependent anoxygenic photosynthesis in the unicellular cyanobacterium *Gloeocapsa alpicola* CALU 743. Int. J. Photoenergy *4*, 169–173.
45. Vermaas, W.F.J. (2001). Photosynthesis and respiration in cyanobacteria. In Encyclopedia of Life Sciences. http://dx.doi.org/10.1038/npg.els.0001670. http://www.els.net/WileyCDA/ElsArticle/refId-a0001670.html.
46. Rajaram, H., Chaurasia, A.K., and Apte, S.K. (2014). Cyanobacterial heat-shock response: role and regulation of molecular chaperones. Microbiology *160*, 647–658.
47. Jeiranian, H.A., Schalow, B.J., Courcelle, C.T., and Courcelle, J. (2013). Fate of the replisome following arrest by UV-induced DNA damage in *Escherichia coli*. Proc. Natl. Acad. Sci. USA *110*, 11421–11426.
48. Garcia-Pichel, F. (1998). Solar ultraviolet and the evolutionary history of cyanobacteria. Orig. Life Evol. Biosph. *28*, 321–347.
49. Schubert, W.D., Klukas, O., Saenger, W., Witt, H.T., Fromme, P., and Krauss, N. (1998). A common ancestor for oxygenic and anoxygenic photosynthetic systems: a comparison based on the structural model of photosystem I. J. Mol. Biol. *280*, 297–314.
50. Anbar, A.D., Duan, Y., Lyons, T.W., Arnold, G.L., Kendall, B., Creaser, R.A., Kaufman, A.J., Gordon, G.W., Scott, C., Garvin, J., and Buick, R. (2007). A whiff of oxygen before the great oxidation event? Science *317*, 1903–1906.
51. Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., Goodrich, J.K., Bell, J.T., Spector, T.D., Banfield, J.F., and Ley, R.E. (2013). The human gut and groundwater harbor non-photosynthetic

bacteria belonging to a new candidate phylum sibling to Cyanobacteria. eLife *2*, e01102.
52. Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G., Steen, J.A., Parks, D.H., Tyson, G.W., and Hugenholtz, P. (2014). An expanded genomic representation of the phylum cyanobacteria. Genome Biol. Evol. *6*, 1031–1045.
53. Harel, A., Bromberg, Y., Falkowski, P.G., and Bhattacharya, D. (2014). Evolutionary history of redox metal-binding domains across the tree of life. Proc. Natl. Acad. Sci. USA *111*, 7042–7047.
54. Price, D.C., Chan, C.X., Yoon, H.S., Yang, E.C., Qiu, H., Weber, A.P., Schwacke, R., Gross, J., Blouin, N.A., Lane, C., et al. (2012). *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science *335*, 843–847.